# Proceedings of the 14th European Conference on eGovernment ECEG 2014

## Spiru Haret University Faculty of Legal and Administrative Sciences Brasov Romania

## 12-13 June 2014

Edited by
Alexandru Ionas
Spiru Haret University
Braşov
Romania

Papers have been double-blind peer reviewed before final submission to the conference. Initially, paper abstracts were read and selected by the conference panel for submission as possible papers for the conference.

Many thanks to the reviewers who helped ensure the quality of the full papers.

These Conference Proceedings have been submitted to Thomson ISI for indexing.

Further copies of this book and previous year's proceedings can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-909507-36-4
E-Book ISSN: 2049-1034
Book version ISBN: 978-1-909507-32-6
Book Version ISSN: 2049-1026
CD Version ISBN: 978-1-909507-40-1
CD Version ISSN: 2049-1042

Published by Academic Conferences and Publishing International Limited
Reading
UK
44-118-972-4148
www.academic-publishing.org

# Community Structure in e-Government Hyperlink Networks

**Paul Henman[1], Rob Ackland[2] and Tim Graham[1]**
**[1]The University Of Queensland, Brisbane, Australia**
**[2]Australian National University, Canberra, Australia**
p.henman@uq.edu.au
robert.ackland@anu.edu.au
t.graham1@uq.edu.au

**Abstract:** Over the past decade the World Wide Web has become a core platform for the electronic operation of government. Yet the shape and nature of government presence on the Web and the online community in which it resides remains poorly understood and relatively under-theorised. This paper is part of a larger project that utilises large-scale web crawling to map the hyperlink network structure between government websites and the broader Web ecology in the UK and Australia. In this paper we utilise Infomap—a state-of-the-art community detection algorithm—to discover 'communities' of websites within a hyperlink network of over 100,000 websites and over 280,000 hyperlinks derived from 88 key UK government seed sites at national, regional (i.e. Scotland and Wales) and local government levels. The principle underpinning the Infomap approach is that flows of information in complex networks reveals community structure. The purposes of analysing online communities in which government websites reside is to identify the different communities operating in this larger network and understand the shared basis for these communities. It is hypothesized that online 'communities' can occur around different policy topics (such as health, education or policing), or along institutional or jurisdictional boundaries (such as England, Scotland and Wales). This paper addresses three main research inquiries. Firstly, what is the nature of the different communities identified in the UK network by the Infomap algorithm, including what types of websites are dominant in each community? Secondly, what role do *government websites* play in each community and what types of sites are dominant in them? Finally, to what extent are government websites included in the *most important* communities. Using this novel approach we examine the extent to which government websites are embedded within the *most important flows* of information on the Web. This research provides foundational knowledge about the role of government websites in the World Wide Web, and the associations that have emerged, and the changing dynamic of state information in the twenty-first century. The research may also lead to new strategies for developing government presence on the Web. Preliminary findings suggest that the social media and government seed sites and portals are key players in the network, though there is considerable diversity in their significance and presence based on policy domain and tier of government.

**Keywords:** social network analysis, e-Government, community detection, hyperlink networks, web social science, UK

## 1. Introduction: E-Government in online networks

The rapid emergence and evolution of the internet since the 1990s generated a much needed academic interest in government use of digital information and communication technologies (ICTs). While the term e-Government is used to embrace the use of digital ICTs by government for political, policy, administrative, and governance activities (e.g. 6, 2004), a great deal of e-Government research has focused on the role of internet (see Henman, 2013). Indeed, key developments in e-Government research has involved a conceptualization and analysis of websites, including the now classical notion of stages of e-Government and the semi-regular world rankings of e-Government by the World Bank and others.

Yet despite this substantial body of research on government on the web, very little is known about the nature of government websites vis-à-vis other websites. In short, much of the analysis of government websites treats the sites themselves as individual items for investigation. This perspective is somewhat strange given that a primary and arguably definitive characteristic of websites is that they are simply an ordered network of webpages within a wider network of webpages and websites joined by hyperlinks. Focusing e-Government analysis on the network in which government websites, instead of the individual sites raises a number of new possibilities. What is the structure of the network in which government websites are located in? What role do these websites take? How easy is it to find government websites for things like public sector information, laws and policies, government decisions, accessing public services, and initiating complaint and appeal processes, and is this information also found in or through other websites? How does the online structure of government relate to offline government as defined constitutionally, organizationally or topic-wise? Indeed, there is already a small body of research examining different elements of these concerns (see for example, Escher et al, 2006; Whalen, 2011; Henman et al, 2014).

This paper seeks to contribute to this knowledge by examining the different online 'communities' in which government websites are located and their role in those communities. We use the word 'communities' in this space carefully to relate to groups of websites highly connected by hyperlinks. As such the relationships are electronically defined, and not by social, legal or organizational relationships. It is hypothesized that online 'communities' can occur around different policy topics (such as health, education or policing), or along institutional or jurisdictional boundaries (such as national, provincial, local).

This paper addresses three main research inquiries based on a network surrounding a selection of key British national, regional and local government websites. Firstly, we seek to understand the different communities identified in the UK network, including what types of websites are dominant in each community. Secondly, we examine the role *government websites* play in each community and the types of government website that are dominant in them (categorised by policy domain and tier of government). Finally, we identify the extent to which government websites are included in the *most important* communities. The remainder of the paper is structured as follows. First, we summarise the various algorithmic approaches to community detection that are available, and explain the nature and rationale for using the Infomap community detection algorithm. The following section explains the research methodology, including the underlying dataset used. The third section then reports the research findings in terms of the structure of the online network vis-à-vis online communities, the nature of those communities and the presence and role of government websites in individual communities and in the community network. The paper ends with a reflection on what this might mean for e-Government research and government website development.

## 2. Networks and 'community detection'

A vast array of natural and digital systems can be represented topologically as networks and examined to identify underlying, and often unforeseen, dynamics and structures. In this study we analyse a *hyperlink network* of UK government websites and the websites connecting to them, hereby referred to as the *UK e-Gov network*. In the UK e-Gov network, *nodes* represent websites and *edges* represent a hyperlink from one website to another. Hyperlinks function similarly to citations, meaning that if a website has an 'outlink' to another website this implies that the other website contains something of value (i.e. information). Thus a website with many 'inlinks' from other websites is generally regarded as authoritative or important. Web users are more likely to 'surf' over to such websites because there are multiple pathways provided by the hyperlink structure of the network as a whole (i.e. many other websites linking to a single website). The size and complexity of the *UK e-Gov* network makes it difficult to examine and analyse effectively. In order to address this problem, we turn to 'community detection'. In general, the aim of community detection is to reduce complex networks into modules that "simplify and highlight the underlying structures and the relationships that they depict" (Rosvall and Bergstrom, 2008, p. 1118).

There are a number of different approaches to 'community detection' in complex networks, including modularity maximization (Lancichinetti & Fortunato, 2012), Edge-Betweenness (Girvan & Newman, 2001), Fast-Greedy (Clauset et al, 2004), Multi-level (Blondel et al, 2008), Walktrap (Pons & Latapy, 2005), and Infomap (Rosvall, Axelsson & Bergstrom, 2009). However, only two of these approaches support the analysis of graphs that are both *directed* and *weighted*. These approaches were Edge-Betweenness and Infomap. In our network under examination, the direction of edges represents the *source* and *receiver* of a hyperlink. That is, whether a website *links to* another website (and vice versa). Similarly, the weight of edges in our network provides information about *how many* hyperlinks a website has to another website. Whereas many community detection algorithms tend to function over the 'underlying' graph (i.e. disregarding the directionality of edges), we require an approach that takes into account both the direction and weight of edges. We utilised Infomap, a cutting-edge approach to community detection in networks. Infomap not only supports directed and weighted networks, but also scales well.

### 2.1 The Infomap approach: Community detection in hyperlink networks

Infomap is an information theoretic approach to detecting community structure in complex networks. Community detection decomposes networks into 'modules' (or communities) by exploiting regularities in network structure. Infomap approaches this task by seeking to describe random walks on networks, taking advantage of their modular structures. For example, we can imagine a 'walker' who is placed onto a node in the network and proceeds to walk randomly from node to node. However, the random walker can only walk to neighbouring nodes via a *directed edge*; the directedness of edges constrains movement because each edge

provides a one-way path, or hyperlink, between neighbouring nodes which may or may not be reciprocated. Similarly, the random walker has a higher probability of walking to a node if the edge has a higher *weight* (e.g. if a website contains 100 out-links to another website rather than just 1 or 2 out-links). These random walks, or flows, are then described by a code that seeks to maximally compress the description of network flow that reveals groups of nodes "among which information flows quickly and easily" compared with the rest of the network (Rosvall and Bergstrom, 2008, p. 1118). These groups of nodes are defined as modules or communities within which the random walker spends a relatively long time before exiting. In short, the central premise of Infomap is that flows of information in complex networks reveal community structure. This enables us to "focus on how the structure of the extant network constrains the dynamics that can occur on that network" (Rosvall and Bergstrom, 2009, p. 14).

The Infomap algorithm returns a list of communities, where each community is a list of nodes (in our case websites). The communities are ordered according to the total amount of flow between the websites in a community as a percentage of total flow between websites in the original full network. Therefore, while communities with large numbers of members (i.e. websites) may be expected to generate a larger proportion of total flow, a highly dense community with fewer websites could account for a greater proportion of total network flow. Infomap also lists nodes (i.e. websites) within each community from highest to lowest flow volume within that community, on the supposition that those nodes with greater flow volume are more important in defining the nature of the network. The Infomap algorithm also returns a list of edges between the communities weighted by flow volume between each community.

## 3. Method

### 3.1 Generating the network data

The sheer size of government online makes is difficult to examine the entire network of government on the Web and the network in which it is located. Moreover, the structure of government on the Web would be expected to be different for different government jurisdictions and policy areas, including between countries. Consequently, the research presented in this paper focuses on government websites in the UK. It extends previous work which focused only on national government agency websites (e.g. Escher et al, 2006; Whalen, 2011) to include regional government (specifically, Scotland and Wales) and local government, as well as the Greater London Authority. Within England, Scotland and Wales, the websites of three rural and three urban local government authorities were included. A selection of websites from key government agencies covering a diverse range of policy and public service areas were identified as reflective of British government on the web. Specifically, sites from central government included the Prime Minister's, Treasury and Parliamentary sites, and the six policy domains: foreign affairs/defence; health; community services (social security, housing); education; environment; law/policing. Government portals were also included. In total, 88 British government websites were identified as seed sites with which to generate a larger network of websites.

Using these 88 websites a large network was created in September 2012 by identifying both all hyperlinks out of each seed website and all hyperlinks coming into each seed site. The hyperlink relationships between these external websites were also collected. Hyperlink network data were collected and assembled via an iterative process that broadly occurred in two stages. Stage One involved webcrawling each seed site using VOSON. This web crawl collected both 'outbound' links (web pages that the seed sites *link out to*) and 'inbound' links (web pages that *link in to* the seed sites). Inbound links were collected via the Blekko API. Further, the internal links for each of the seed sites were collected (to a maximum of 1500 pages). Stage Two involved finding the outbound links for pages that were discovered during Stage One (these pages are known as the 'first ring'). Hyperlinks between pages in the first ring were mapped during this process, resulting in a network of nodes (i.e. web pages). Next, 'pagegroups' were created from pages in the first ring, to enable network analysis between websites (i.e. domain names) rather than only between individual webpages. Thus, each node in the network is a collection or grouping of webpages (generally a single domain name). This resulted in one network, where each node represents a separate website/domain name, containing over 100,000 websites and over 280,000 hyperlinks.

### 3.2 Generating 'communities' using Infomap

The Infomap algorithm was then applied to this very large hyperlink network. The Infomap algorithm allows various parameters to be specified which may affect the output. We used two parameters: (1) specifying that

the network is *directed*; and (2) specifying N, the number of attempts to partition the network. The latter parameter could significantly affect the results. However, we found only minor differences to output results for values of N greater than 10, and no differences for N greater than 100. We report results for N=100.

## 4. Results: The community structure of the UK e-Gov network

### 4.1 Enumerating and naming communities

The Infomap algorithm generated 4571 unique communities from the UK e-Gov network, and resulted in 280,580 links *between* communities. The median number of websites per community is 126.5. The maximum number of websites in a single community is 1830 (Community 3 – '*Blogosphere (Blogger)*') and there are 1018 communities that contain three or less websites.

A first observation in the communities derived from the analysis is that a relatively small number of communities summarise or make up much of the entire network. Although 4571 communities were generated, a very small percentage of communities accounted for most of the overall flow in the full network. Indeed, the distribution of flow between communities in the UK e-Gov network follows a power law distribution. This is often referred to as a 'scale free network' (Barabasi & Bonabeau, 2003), which has been observed more broadly on the Web, whereby a relatively small number of popular websites account for a majority of the world's internet traffic. While the majority of communities share only a small portion of flow across the network, a very small handful of communities account for the majority of flow. In the UK e-Gov network, the top 1% (50 nodes out of the total 4571) account for approximately 72% of all flow. Similarly, the top 400 (11.4%) nodes account for over 90% of all flow. Table 1 provides a list of the Top 25 communities, presents the communities in descending order by 'flow' volume between communities (loosely interpreted as 'importance'), and also lists the number of websites (or nodes) from the entire network in each community.

A second component in the analysis was to understand the nature of each community, in terms of any shared characteristic within each community. While Infomap automatically labels communities after the website that has the highest flow volume in each unique community, this can be a useful heuristic, but it also can be misleading. Thus, identifying how to meaningfully label each community is both critically important as a first analysis step, and also to build further analysis upon that. Naming helps inform us about 'what is going on' in each community. As Rosvall and Bergstrom (2008, p. 1118) write: "useful maps assign unique names to important structures". In practice we observed that a considerable number of communities contain a heterogeneous assortment of websites that do not readily suggest a clear category or label. Although the automatic label (i.e. the website with highest flow) provides a clue to what is occurring in each particular community, it can also be potentially misleading. For example, Community #20 in Table 1 contains a high proportion of websites specifically relating to Edinburgh, but also a high proportion of websites relating to Scottish Education more broadly. Although it is not yet clear exactly why these two domains co-constitute a single community, our analysis suggests that communities do not always simply revolve around the top ranking website (e.g. www.edinburgh.gov.uk in Community #20). In this case we named Community #20 as *'Edinburgh/Scottish Education'* rather than just '*Edinburgh*'. Therefore, naming communities requires a fair degree of qualitative decision making. The rationale needs to take into account website statistics (e.g. by policy domain, tier of government, generic top level domain) as well as critical awareness of the role and dynamics of key websites and how they relate to the rest of the community. As explained below, further analysis of the composition of each community by other characteristics were undertaken.

Based on this naming convention, Figure 1 is a network visualisation of the resulting communities in the UK e-Gov network. It displays the *links* with the highest flow in the network and the 33 communities attached to them (accounting for over 60% of all flow in the network). The size of a community (the 'circles') is proportional to the average time a random walker spends on websites in the community. The width of a link (the lines between communities) is proportional to the per step probability that a random walker moves between the communities.

In Figure 1 we observe that the community named *Twittersphere* (i.e. www.twitter.com and the community forming around it) has an important and central role in the overall community structure. *Twittersphere* has the highest flow out of all communities in the UK e-Gov network and is therefore the No. 1 Community according to Infomap. It appears to be a central 'hub' connecting many communities that otherwise have weak or non-

existent links. Furthermore, the *Twittersphere* directs a massive amount of flow to the *Parliament* community (i.e. www.parliament.uk and the community forming around it). This suggests that the *Twittersphere* community plays a critical role in routing information and traffic to UK parliament on the Web.

**Table 1**: Top 25 communities (descending order by flow volume)

| Rank | Community | Aggregated flow volume | Size (number of websites) |
|---|---|---|---|
| 1 | Twittersphere | 0.0746714 | 356 |
| 2 | Parliament | 0.0507411 | 148 |
| 3 | Blogosphere (Blogger) | 0.0476225 | 1830 |
| 4 | Community Sector | 0.044557 | 509 |
| 5 | Wales | 0.0353763 | 152 |
| 6 | Health | 0.0302951 | 1755 |
| 7 | Google / YouTube | 0.0282725 | 271 |
| 8 | UK Central Government | 0.0249578 | 326 |
| 9 | Scotland | 0.0237664 | 1063 |
| 10 | Bermuda | 0.0198672 | 45 |
| 11 | News (and news aggregation) | 0.019325 | 531 |
| 12 | Environment | 0.0182983 | 466 |
| 13 | Policing | 0.0172842 | 1126 |
| 14 | Business | 0.0166265 | 97 |
| 15 | Media and data sharing | 0.0162993 | 190 |
| 16 | Facebook | 0.0151511 | 1 |
| 17 | London | 0.0139016 | 673 |
| 18 | Open Gov / Open Data | 0.0135712 | 106 |
| 19 | Information Reference (e.g. Wikipedia) | 0.0131204 | 711 |
| 20 | Edinburgh / Scottish Education | 0.0122874 | 694 |
| 21 | Recycling and Sustainability | 0.0120381 | 701 |
| 22 | Global Health | 0.0104746 | 148 |
| 23 | Europe / International | 0.0101316 | 708 |
| 24 | Blogosphere (Wordpress) | 0.00978866 | 1320 |
| 25 | Scottish Politics / Ombudsman | 0.00901493 | 455 |

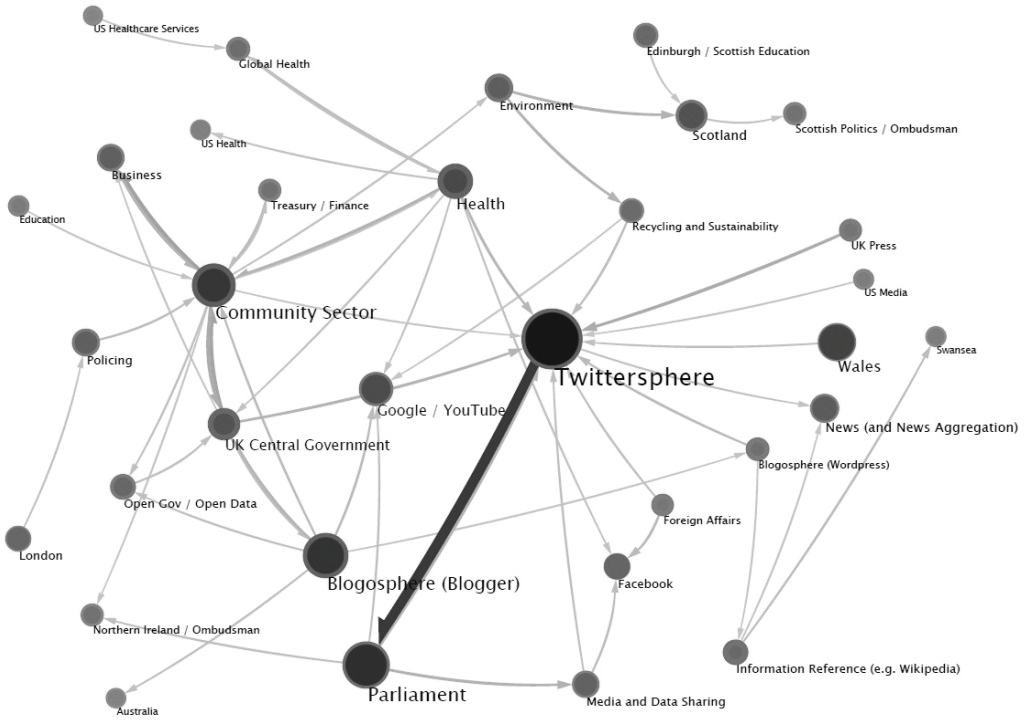


**Figure 1:** Visualizing the community structure of UK government on the Web (by *link flow*)

## 4.2 The makeup of website communities

We also sought to quantitatively understand the makeup of significant communities. Table 2 provides a breakdown of the Top 10 communities by *Country Code Top Level Domain* (i.e. uk, au) and *Generic Top Level Domain* (e.g. gov, com, org). We observe that the communities are dominated by UK websites or unknown country code, the latter of which are sites ending in .com, .org and so on. This makes sense given the seed sites are British. Significantly, social media and blogging communities (i.e. Communities #1, #3 and #7) have majority websites unknown and about 20% UK websites. Whereas communities with a focus on government (i.e. #2, #8), a geographical domain (i.e. #5, #9) or a policy area (i.e. #4, #6) are made up of between a half to two-thirds UK websites. The notable exception is Community #10 ('Bermuda'). In this community, 69% of websites are exclusively from Bermuda (i.e. *bm*), representing a significant degree of homogeneity compared with other communities.

**Table 2**: Breakdown of all websites in Top 10 communities by country code and web domain

| | Community | Country Code TLD | | | Generic TLD | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | UK | Other | Unknown (e.g. com, org) | Gov | Org | Net | Com | Other |
| 1 | Twittersphere | 20% | 20% | 60% | 2% | 16% | 5% | 63% | 16% |
| 2 | Parliament | 47% | 5% | 48% | 14% | 33% | 4% | 37% | 13% |
| 3 | Blogosphere (blogger) | 23% | 14% | 63% | 1% | 20% | 4% | 63% | 12% |
| 4 | Community Sector | 67% | 3% | 30% | 17% | 34% | 2% | 43% | 5% |
| 5 | Wales | 63% | 2% | 36% | 12% | 48% | 3% | 35% | 3% |
| 6 | Health | 62% | 3% | 35% | 2% | 39% | 2% | 51% | 6% |
| 7 | Google / YouTube | 19% | 31% | 50% | 3% | 30% | 3% | 36% | 29% |
| 8 | UK Central Govt | 50% | 8% | 42% | 21% | 29% | 3% | 39% | 9% |
| 9 | Scotland | 56% | 3% | 41% | 7% | 35% | 3% | 49% | 7% |
| 10 | Bermuda | 5% | 69%* "bm" | 27% (*other+ unknown*) | 45% | 7% | 2% | 9% | 38% |

*** For *'Bermuda' ,* 69% of websites are from Bermuda ('bm')

In terms of Generic Top Level Domain, government websites (i.e. gov) do not constitute a significant proportion of any community, except for Bermuda, where it constitutes 45% of all websites in that community. Rather, websites in the commercial (.com/.co) or organizational (.org) domain make up the large bulk of each community, which is not surprising given global statistics of domain registrations. Government websites tend to account for a substantial proportion in Community Sector (17%), Parliament (13.5%), and UK Central Govt (20.5%). The low proportion of government websites in Health may be surprising until we recognize that health is a domain in which a lot of private organisations operate alongside of non-government/non-profit health-related agencies for service delivery and consumer advocacy and the like.

**Table 3**: Government websites in Top 10 communities by Tier of Government/Jurisdictional Boundary

| Community | National | Greater London | Ireland | Scotland | Wales | Local |
|---|---|---|---|---|---|---|
| Twittersphere | 1 | 0 | 0 | 0 | 0 | 1 |
| Parliament | 6 | 0 | 0 | 0 | 0 | 0 |
| Blogosphere (blogger) | 2 | 0 | 0 | 0 | 0 | 0 |
| Community Sector | 36 | 1 | 1 | 3 | 4 | 9 |
| Wales | 0 | 0 | 0 | 0 | 20 | 0 |
| Health | 17 | 3 | 0 | 1 | 0 | 5 |
| Google / YouTube | 3 | 1 | 0 | 0 | 0 | 0 |
| UK Central Govt | 24 | 0 | 0 | 1 | 0 | 1 |
| Scotland | 1 | 2 | 0 | 48 | 0 | 30 |
| Bermuda | 0 | 0 | 0 | 0 | 0 | 0 |

For the British government websites within communities, we are also able to identify what tier of government they are associated with. Table 3 provides a breakdown of the Top 10 communities by Tier of Government/Jurisdictional Boundary (e.g. Scotland, Wales, Local, National). We observe that some communities seemed to exhibit fairly strict jurisdictional boundaries. For example: Parliament contained only

'National' websites (100%); Wales contained only 'Wales' websites (100%); and UK Central Govt almost exclusively contained 'National' websites (approximately 92%). In contrast, we also found that some communities challenged jurisdictional boundaries. For example: Community Sector contained websites across all 6 categories (although predominantly 'National' and 'Local'); and Scotland contained a significant proportion of 'Local' as well as 'Scotland' websites, which contrasted with its counterpart Wales (which contained zero local websites). Overall, there appears to be a logic to these results.

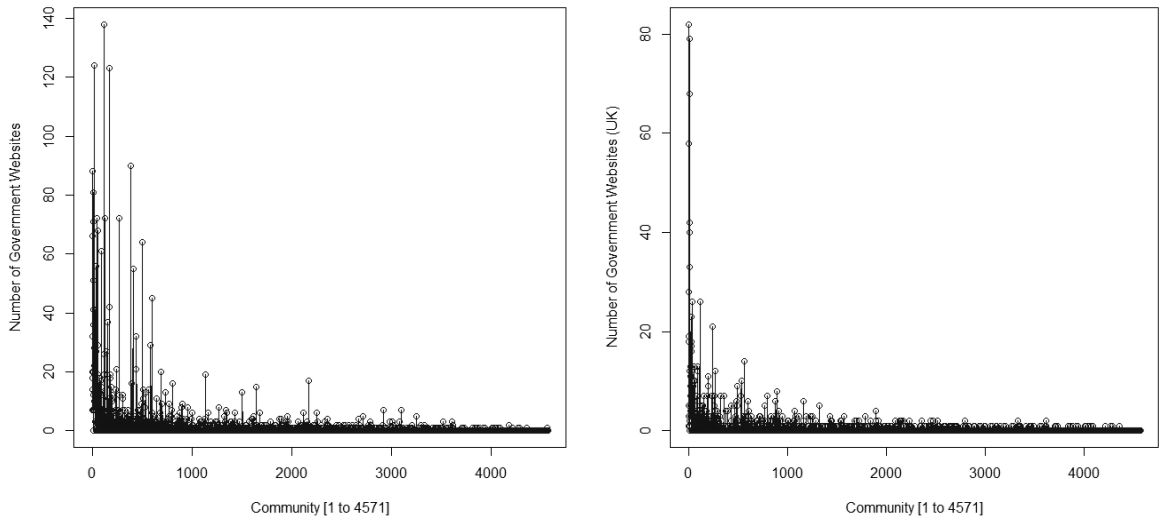


**Figure 2:** Number of government websites in each community. 2a (LHS) all government; 2b (RHS) UK government

Looking across all the communities, we also sought to understand the extent to which government websites are part of communities, and more specifically in the *most important* communities. There are a total of 4287 government websites in the e-Gov UK network (out of a total 107462 websites). These are defined as any website with a Generic Top Level Domain ending with 'gov', 'go', 'govt' or 'gv', which includes other nation's government websites (i.e. not just UK). The total of 4287 government websites includes 1679 *UK government* websites (i.e. the *gov.uk* domain). Figure 2(a) illustrates the spread of all 4287 international government websites (y-axis) across all 4571 communities (x-axis), while Figure 2(b) plots only UK government websites (i.e. 1679 websites in the *gov.uk* domain). Recall that communities are ordered in terms of importance in ascending order (i.e. Community 1 is most important). Across both figures we observe a strong pattern of government websites towards the left-hand side of the graph. This indicates that, in general, both international and UK government websites are included in the most important communities in the UK e-Gov Web ecology.

## 5. Conclusion

In this paper we provided foundational knowledge about the community structure in which UK government websites are located. We discovered that on the basis of 88 British government websites from national, regional and local levels across a range of policy domains, generated a wider network of over 100,000 websites. Within this network Infomap generated many thousands of communities, yet only a small percentage of these communities are of any significance; out of 4571 communities, the top 1% account for over 75% of all flow throughout the network. Moreover, we discovered that the types of websites within each community are generally quite diverse, including a mix of Generic Top Level Domains and Country Code Top Level Domains, yet there also appeared to be a theme or two within the sites (see Table 1). We observed that government websites within a community exhibit diverse combinations of tiers of government and policy domains. However, we also noted a small number of tightly-knit communities that exhibit fairly concise jurisdictional or policy boundaries (e.g. *Health*). Furthermore, government websites tend to be included in the most important communities of the large network, based on flow volume. This applies to both UK government (i.e. gov.uk) and non-UK government websites.

Our findings indicate that the Infomap algorithm is highly suited to this kind of research and provides a useful conceptual tool to examine the 'social life' of large-scale Web networks. Our initial expectations about the structure of government on the Web are partly supported by the results, which reveal that to some extent communities do form around different policy topics and institutional or jurisdictional boundaries. However, there are at least two caveats to this hypothesis. Firstly, such communities are generally quite diverse in terms of their constituent websites (i.e. they often transgress institutional boundaries of policy domain and jurisdiction). Secondly, whilst important communities formed around government seed sites, we also found that these communities competed equally for visibility and dominance with non-government communities such as *Twittersphere* (#1), *Blogosphere* (#3), *Google/Youtube* (#7), *News and News Aggregation* (#11), and *Business* (#14). Furthermore, communities included a mix of both government (gov) and non-government websites (org, net, com), challenging the notion that government on the Web is structured similarly to traditional 'offline' arrangements. Indeed, government sites never make up a large component of websites in a community, though this is perhaps not surprising, given the makeup of the World Wide Web, and the population of offline organisations.

Given that 8 out of the top 20 communities are not government orientated, it is crucial that e-Government schemes continue to forge strategic links and relationships across the commercial and non-profit Web. In particular, social media and blogging communities are extremely important players in the overall Web ecology. Governments will benefit from building in-roads and out-roads within, and between, these communities. Moreover, the success of e-Government Web strategies can be quantified by measuring the extent to which government websites are positioned in terms of community structure. That is, whether government websites are strategically networked at the center of information flows on the Web, which we have shown can be examined and visualized in terms of communities.

To sum up, we offer two points for consideration. Firstly, governments are continually expanding and evolving Web presence in order to improve and optimize e-Government projects. Community detection of government hyperlink networks provides a useful and novel tool for strategic e-Government analysis and management toolbox. Analysing community structure provides a unique window into the changing nature of state information in the 21st century; governments can examine the Web-networked relationships between various arms of the state and the broader Web ecology and act on this knowledge to repair, create and strengthen linkages. Finally, UK and other governments have moved to centralise existing government websites into 'one-stop shop' portals (e.g. www.gov.uk). Our data suggests that this has affected the community structure of government on the Web because nodes in the network (i.e. separate government websites) collapse into one super-node (i.e. a central portal website). We currently do not know whether, and how, this affects the flow of information in such networks. Given that our web crawling data were collected in late 2012, future studies might examine what kinds of community structures have since emerged from the 'portalisation' of government websites and the implications for e-Government.

## Acknowledgements

We gratefully acknowledge the support of the Australian Research Council Discovery Grant (DP110100446), and the technical computer expertise contributed by Lin Chen in the first half of this project. We also gratefully acknowledge the insightful and technical assistance from Martin Rosvall (InfoMap creator) in successfully utilising InfoMap.

## References

6, P. (2004). *E-governance: styles of political judgment in the information age polity*, Basingstoke: Palgrave.

Ackland, R. (2011). WWW Hyperlink Networks. In Hansen, D. L., Schneiderman, B., & Smith, M. (2011, Chapter 12). Analyzing social media networks with NodeXL: Insights from a connected world. Amsterdam: M. Kaufmann.

Barabasi, A., & Bonabeau, E. (2003). Scale-free networks. New York: Sci American Inc.

Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 70(6 Pt 2), 066111.

Edler, D. & Rosvall, M. (2013), The MapEquation software package, available online at http://www.mapequation.org.

Escher, T. & Margetts, H. (2006). Governing from the Centre? Comparing the Nodality of Digital Governments. Available at SSRN: http://ssrn.com/abstract=1755762

Girvan, M., & Newman, M. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 99(12), 7821-7826.

Henman, P. (2013). Government and the Internet: Evolving Technologies, Enduring Research Themes. In Dutton, W. H. (2013, Chapter 14). The oxford handbook of internet studies. Oxford: Oxford University Press.

Henman, P., Ackland, R. & Graham, T. (2014) Social Policy on the Web: The Online Institutional Structure of Social Policy Domains in the UK

Lambiotte, R., & Rosvall, M. (2012). Ranking and clustering of nodes in networks with smart teleportation. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 85(5 Pt 2), 056107.

Lancichinetti, A., & Fortunato, S. (2011). Limits of modularity maximization in community detection. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 84(6 Pt 2), 066122.

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. (pp. 284-293). Berlin: Springer-Verlag Berlin.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences of the United States of America, 105(4), 1118-1123.

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. The European Physical Journal Special Topics, 178(1), 13-23.

Whalen, R. (2011). The Structure of Federal eGovernment: Using hyperlinks to analyse the .gov domain. Paper 8. Available at: http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1009&context=pnconfs_2011